

Understanding the Mood of Social Media Messages

Robert Power

CSIRO

robert.power@data61.csiro.au

Bella Robinson

CSIRO

bella.robinson@data61.csiro.au

Amanda Dennett

Australian War Memorial

amanda.dennett@awm.gov.au

Brian Jin

CSIRO

brian.jin@data61.csiro.au

Cecile Paris

CSIRO

cecile.paris@data61.csiro.au

Abstract

Social media is a valuable source of information when seeking to understand community opinion and sentiment about issues of public interest. Such analysis is usually based on sentiment or emotion processing using machine learning techniques or references a curated lexicon of words to measure the emotive intensity being expressed. The lexicon approach can be limited by the sparsity problem, where the lexicon words are not present in the text being processed, and context issues, where the lexicon words have different meanings in the domain under investigation. We have developed a novel technique based on word embeddings to mitigate these issues and present a case study showing its application, where the mood expressed by the community on social media about the Centenary of Armistice in Australia was determined in near real-time.

1. Introduction

Social media provides citizens with a public platform to discuss many topics and voice their opinions about them. Mining these discussions can provide insights about a variety of topics of interest to government agencies, business enterprises and non-profit organisations [19,30]. These insights can then be used for a variety of purposes, such as designing better information campaigns, understanding concerns and issues of relevance to citizens, the development of new services and products, improved service delivery, marketing and brand management [26,32].

A key aspect of this social media processing is the ability to classify messages to identify the underlying sentiment or emotion being expressed. There are subtle but important differences between sentiment and emotion. From a psychological perspective, emotions are an expression of a complex personal state involving multiple layers influenced by various factors such as

belief, experience, background and culture while sentiment is a subjective opinion about a topic [20].

In terms of natural language processing (NLP), these differences are seen by the processing outputs resulting from these two tasks. Sentiment analysis produces a ternary positive/negative/neutral result whereas emotion analysis uses different labels, such as: love, joy, anger, sadness, fear and surprise [24]. While there is no standard agreed list of emotion categories, recent research suggests there are four primary emotions of anger, fear, sadness and happiness with more complex secondary emotions being derived as combinations of these primary ones [29].

These categories and founded in the sociology of human emotions and using them as labels for NLP analysis may not reflect the insights under investigation for a social media campaign or opinion mining task about a specific topic or event: the domain under investigation. For this reason, we have adopted a domain specific list of categories that correspond to the different feelings and opinions expected to be expressed on social media for our event of interest. This is also why we have adopted the term ‘mood’ rather than emotion: to emphasise that we are undertaking a classification task related to the well-known emotion analysis, but which uses a new target set of labels that is specific to our study. Our methods should be applicable to others who undertake social media analysis to gain insights on topics that may not align with the categories commonly used for emotion analysis.

The rest of the paper is organised as follows. First, we describe previous work on sentiment and emotion analysis in social media. Then we present our problem statement as a case study which defines the motivation for this work. Our solution is then described, a combination of a rules based approach and the application of word embeddings as an alternate classification technique. Our experiments and the results are then presented and the paper finishes with our conclusions and plans for future work.

2. Related Work

Social Media is a valuable source of consumer views and public opinion which is of interest to business, governments and the community. Making use of this data is facilitated through sentiment and emotion analysis which identifies and measures opinions and feelings from text. Achieving this on social media faces numerous challenges such as the use of informal language, abbreviations, lack of training data, the ambiguity of language, the presence of sarcasm and so on [5,22,30,32].

The NLP task of emotion analysis relies upon a set of descriptive labels to produce a result however there is no single standard agreement of emotions. Ortony and Turner [20] compared 14 different lists of emotion categories to explore the notion that there exists a small set of ‘basic’ emotions. These lists vary in the number of emotions included, with two lists containing only two emotions, pain/pleasure and happiness/sadness, while another list has 11: anger, aversion, courage, dejection, desire, despair, fear, hate, hope, love and sadness. While they concluded that there is no criterion to define ‘basic’ in this field of study, subsequent work has suggested that there are four primary emotions of anger, fear, sadness and happiness which are universal across cultures and more complex secondary emotions are derived as combinations of the primary ones [29].

Emotion analysis of social media has been studied to identify hate speech [14,23], monitor corporate sentiment and brand management [26,31], understand issues of depression [8,18], monitor happiness [9,10,12] and calculate emotion intensity [4,15,27]. Lexicons of emotion words have been established [3,5,9,18,25,32] to help with this analysis.

SentiStrength [27] estimates the strength of positive and negative sentiment in short texts. It has been applied to Twitter data streams measuring ‘Gross National Happiness’ using a Tweet ‘polarity of the day’ metric [10]. It has also been used to explore question/answering systems, blog comments and issues of gender bias in online reviews [28].

Sentiment intensity has been proposed as an alternate mechanism to identify events in social media streams. Rather than looking for a burst of activity in terms of tweet volume, the strength of sentiment in the messages can be used. This approach has been used to automatically identify emergency events from tweets based on sentiment classification [7].

Issues related to corporate brand management are supported through sentiment analysis techniques. Subtle sentiment expressions in Twitter were shown to be useful in a case study involving public comments about Starbucks [31]. Similarly, the corporate crisis involving the Volkswagen ‘Dieselgate’ emission

scandal showed there was a strong public reaction to the scandal over a six-week period [26].

The use of emojis for emotion status monitoring has also been studied [6]. They found that including emojis as part of the analysis improves the sentiment result, achieving state-of-the-art performance. Others have found the use of emojis convey different meanings in different contexts [13].

The task of analysing the sentiment of text is achieved either using lexicon-based approaches or supervised learning methods [5]. Lexicon based approaches [3,9,18,25,32] are useful since they can readily be applied to tasks without the need to gather and label training data and they are flexible in terms of the ability to control the lexicon vocabulary. However, their drawbacks are the lack of contextual information, potential disjoint vocabularies or nuanced indicators of sentiment expression [32]. We refer to these issues as the context problem: the lexicon words need to consider the situation being investigated.

The application of lexicon-based approaches for tweets may also suffer from the sparsity problem: what to do when none of the lexicon words are present in the tweet text? This issue is addressed in the Hedonometer tool [9] by using a large lexicon of over 10,000 words and in the We Feel tool [18] by processing a large volume of tweets. These systems provided motivation for our work and so are described in more detail next.

Hedonometer [9] provides a Twitter happiness measure based on a large corpus of words ranked on a nine-point happiness scale from 1 (sad) to 9 (happy). There are over 10,000 frequently occurring words that were ranked using Amazon’s Mechanical Turk service. Their website, <https://hedonometer.org/>, shows the daily variation in the happiness of tweets from the Twitter sample stream by averaging the happiness scores of the Hedonometer words found in the tweets. They also note that their corpus on the happy/sad spectrum can be extended to other emotion words, such as fear, anger, and surprise, although there does not appear to have been any recent work in this area.

We Feel is a similar tool which looks at whether the emotions of the world can be mapped in real-time from tweets. It also examines if indicators of mental health issues can be measured globally from this data source [18]. The We Feel tool, <http://wefeel.csiro.au/>, presents a real-time dynamic visualisation of emotions extracted from tweets. Like Hedonometer, We Feel uses a corpus of emotion bearing words, however they have organised their words into a hierarchy structured as primary, secondary and raw emotions. There are 532 ‘raw’ emotion words which were categorized using a crowdsourcing task against Parrott’s hierarchy of emotions: love, joy, surprise, anger, sadness, and fear [21].

3. Motivation

3.1. Case Study

During late 2018, the Australian War Memorial (AWM) commemorated the Centenary of Armistice over a five-week period from 5 October to Remembrance Day, 11 November. The theme was 'Honour Their Spirit', and several public activities were held, notably 62,000 hand-crafted red poppies on the Memorial's grounds representing the Australian lives lost in the First World War.

Digital engagement activities were also delivered by the Memorial and the public were encouraged to share their experiences and thoughts about the centenary on social media. The Memorial actively promoted the event using their social media accounts [2], adopting the hashtag #HonourTheirSpirit. The AWM are active on social media and used the Stackla tool, <https://stackla.com/>, amongst others, to help manage this campaign and public engagement.

3.2. Problem Statement

The main objective of the AWM was community engagement. This was facilitated through an interactive web-based display to present an onscreen visualisation of the mood being expressed by audiences engaging with the Memorial's social media platforms [1]. The main task was to process the social media messages into mood categories that accurately reflect the feelings being expressed by the public.

This was developed in an iterative process. First, relevant social media content was collected in advance of the official commemoration period which provided data for training purposes. Simple analysis of this data, including word frequency counts, topic analysis, extracting key hashtags, phrases and emojis provided an initial collection of candidate rule conditions to categorise the messages.

Our emotion, or mood, analysis was motivated by the lexicon-based approaches similar to Hedonometer and We Feel. We were mindful of the lexicon issues of sparsity and context and overcoming these problems formed the research focus of our investigation.

3.3. Data Collection

Relevant social media data was collected to explore the topics being discussed to establish an initial set of rules for identifying the underlying moods being expressed by the community. This included public comments on the Memorial's social media accounts [2] and queries of the hashtags #australianwarmemorial

and #awmemorial from Twitter and Instagram. Other keyword, phrase and hashtag queries were also used such as: ANZAC, #ANZAC, Remembrance Day, Armistice Day, #RemembranceDay2018. We also collected geo-tagged content posted at the memorial, as Remembrance Day and Armistice Day are also commemorated by other Commonwealth countries.

This data collection began in mid-April just before Anzac Day 2018 which is a national day of remembrance in Australia and New Zealand that originally commemorated the Australian and New Zealand Army Corps (ANZAC) First World War landing at Gallipoli on 25 April 1915. This event provided social media content like that expected during the lead up to Remembrance Day. This data collection continued until 19 November 2018. During this period the AWM delivered a range of public programs to mark the 100th anniversaries of significant battles from the First World War. Visitation to the Memorial was higher than usual during this time, as was public activity related to commemoration on social media. A summary of the volume of data collected is shown in Figure 1 where the peaks correspond to Anzac Day and Remembrance Day. In total, 7,770 Instagram messages, 30,425 Facebook posts and 127,120 tweets were obtained.

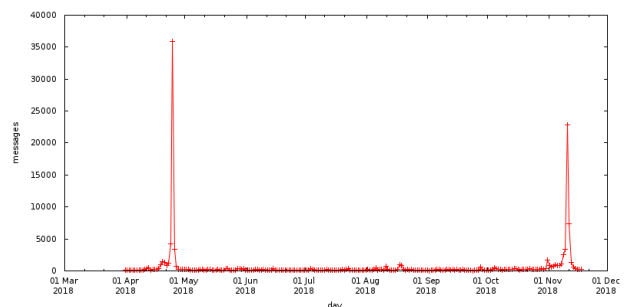


Figure 1: Initial data volume collected per day.

Since there were different social media monitoring strategies undertaken, the data collected was categorised in terms of its expected relevance to the topic of the Centenary of Armistice as a scale of 1-3 where 1 is considered strongly relevant, 2 is relevant and 3 somewhat relevant. These categories are:

1. Content from the Memorial's Stackla API.
2. Public comments on the AWM's Facebook page and Instagram messages and tweets that include specific keywords, such as #honourtheirspirit.
3. Tweets that include an expanded set of keywords such as #lestweforget and #remembranceday and phrases such as 'lest we forget' and so on.

A summary of the number of posts collected per social media platform by category is shown in Table 1. Note there are no Facebook posts for category 1 since

the Stackla API did not provide public comments posted on the Memorial's Facebook page. We elected not to process content from the Memorial since we are interested in public opinion, not the Memorial's.

Table 1: Summary of social media data collected.

Category	Platform	Count
1	Instagram	5,974
	Twitter	1,015
2	Facebook	30,425
	Instagram	1,796
	Twitter	21,554
3	Twitter	104,551

4. Solution Strategy

4.1. Mood Categories and Rule Conditions

The original mood words were proposed by the Memorial staff, based on their experience of issues relevant to the AWM and from the analysis of the initial data collection. These categories were discussed and refined until the final list was agreed as: respect, gratitude, humility, pride, sadness and discontent. The aim of this approach was not to disregard established work on emotion lexicons, but rather to ensure that the public digital display published by the AWM showed the moods in groupings that related directly to remembrance and commemoration, rather than traditional groupings such as joy and anger.

These six mood categories were then expanded to identify associated words, phrases, hashtags and emojis. As an example, the elements associated with the mood 'respect' are shown in Table 2. In total there are 106 words, 12 phrases, 13 hashtags and 43 emojis used; see <http://awm.csiro.au/rules.html> for the full list.

Table 2: Text elements for the mood 'respect'.

Word	Phrase	Hashtag	Emoji
beloved	'armistice day'	#ArmisticeDay	❤️
ceremonies		#DawnService	🙏
ceremony	'dawn service'	#LestWeForget	💜
commemorate		#WeWillRememberThem	💙
commemorating	'lest we forget'		💖
commemoration			💗
commemorations	'we will remember them'		💕
commemorative			💞
memorial			🥰
poignant prayer			😍
reflection respect			💖
shrine silence			💗
solemn spirit			💕
thoughts tradition			💞
wreath			💖

4.2. Rule Processing

Our rule-based approach is a pattern matching task to find the target conditions in the message. These patterns are the hashtags, emojis and phrases that are associated with the six moods. URLs and stopwords are removed from the messages along with platform specific elements, such as retweet indicators and user mentions.

There are a few special cases to be considered in this processing, such as case specific matching of some abbreviations, for example 'RIP' (as an abbreviation for 'Rest In Peace') to distinguish this from the word 'rip', and matching the phrase 'thank you for your service' before checking the phrase 'thank you'. All phrases are also matched twice: initially for the phrase itself, such as 'thank you' or 'lest we forget', then again with the spaces removed: 'thankyou' and 'lestweforget'. All phrase and hashtag matching is case insensitive apart from exceptions such as 'RIP'.

In summary there are four rules applied which examine the following elements of the text message: hashtags; emojis; phrases; and phrases with spaces removed. Our initial processing using the test data collected found that using rules alone we could associate about 60% of the messages to a mood.

The tweet in Figure 2, showing the poppy exhibition on the Memorial grounds, illustrates this process. The tweet text contains three elements that match rule conditions: the phrase 'Honour their Spirit', the hashtag #HonourTheirSpirit and the 'ok hand sign' emoji. The first two rules are associated with the mood 'humility' while the last one corresponds to 'gratitude' with the result that this tweet would be associated with these two moods: the rule processing can assign a single message to multiple moods.

4.3. Word Embeddings

As noted above, the rules can classify about 60% of messages to one or more moods. To increase this percentage, other processing options were investigated. Motivated by the Hedonometer and We Feel tools, we wanted to explore our own curated lexicon of words and chose to investigate word embeddings for this task.

Word embeddings are an NLP term given to a collection of techniques that associate words or phrases to a numeric representation such that the context of the words is maintained in the numeric representation: words semantically 'close' are also numerically close in the model. We chose to use the Google word2vec word embedding toolkit due to its popularity, implementation maturity and toolkit availability [16]. We used the gensim Python package and the Google model 'GoogleNews-vectors-negative300.bin.gz'.

This model can be used to look for words that are close to our lexicon of words from our mood categories. However, the context similarity means words such as ‘love’ and ‘hate’ can be numerically close in the model since they can be used in similar contexts. As an example, the six words closest to ‘love’ in the model are in order: loved, adore, loves, passion, hate and loving. While ‘hate’ is the fifth closest word to ‘love’ the remaining words are all semantically close to the intended meaning of ‘love’ for our purposes. This is also the case for the 106 mood words in our lexicon: examining the 10 closest words for each, there are 13 instances where there are minor semantic conflicts similar to the love/hate example above. Of these, in two cases the closest word in the model has the opposite meaning: remember/forget and fallen/risen. This highlights the need to review the lexicon words before processing and we discuss this issue further in the conclusion.

Honour their Spirit 1918-2018 62,000 killed, a heartfelt tribute and must see at Australian War Memorial 🌹 #HonourTheirSpirit



6:40 PM - 15 Oct 2018 from Australian War Memorial

Figure 2: Example tweet¹.

4.4. Worked Example

In summary, the word2vec model is used by taking each word in the social media message and calculating the cosine similarity distance, the ‘score’, to the 106

mood words associated with the six mood categories. This distance measure has a range of [-1.0–1.0], where -1.0 is far away and 1.0 means the words are the same.

There are options for deciding which is the best mood category for a target word, such as finding the best score in each category or taking the average. These alternatives can be illustrated using the tweet of Figure 2 as a worked example. When the emoji, URL and image are removed, the message from Figure 2 is:

Honour their Spirit 1918-2018 62,000 killed, a heartfelt tribute and must see at Australian War Memorial #HonourTheirSpirit

Although this message can be associated to a mood category using the rules, we are using this text as a worked example to simplify the discussion. This message is parsed to remove elements that match the predefined rule conditions (phrases ‘Honour their Spirit’, #HonourTheirSpirit and the emoji) and stop words are also removed. This leaves the six words that are present in the word2vec model: killed, heartfelt, tribute, Australian, War, Memorial.

The highest cosine similarity distance scores for these six words when compared against the 106 mood words are shown in Table 3. Some notes about the processing:

- So long as there are (non-stop) words in the message present in the model, then a score will be found. It may however not be a ‘good’ score.
- The Google word2vec model contains mostly words with US spelling. When a word from the message is checked to be in the model, if it is not found, then there is another check done to see if a corresponding US spelling for the word exists and if so, this US word is substituted.
- The 106 mood words contain both US and UK spelling. For example, ‘honoring’ and ‘honouring’ are both mood words in our lexicon.

Table 3: Example word2vec results.

Message word	Best mood word	Mood category	Score
killed	killed	sadness	1.00
heartfelt	poignant	respect	0.69
tribute	honoring	pride	0.60
Australian	clash	discontent	0.11
War	commemoration	respect	0.24
Memorial	memorial	respect	0.59

A simple solution to assign a mood to a message is to take the highest score as the result. In this case, ‘sadness’ would be attributed to the message. There are other options for this process which are explored next.

¹ <https://twitter.com/LindaBooBruce/status/1052011484107034625>

5. Experiments

Several experiments were performed to determine the best method of using the cosine similarity distance results to associate a message to one of the mood categories of respect, gratitude, humility, pride, sadness and discontent.

5.1. Word Scoring Options

We introduce the term ‘scoring option’ to refer to the different processing methods to associate text to a mood category. While the goal is to label the whole message to a mood, our initial investigations were to do this for the individual words in the message. Table 3 showed one method for doing this: choose the ‘best’ mood category with the closest associated mood word:

```
find_best_mood_category(word):
    (max, best) = (-2, "")
    foreach mood in lexicon:
        score = model.dist(word, mood)
        if score > max:
            (max, best) = (score, mood)
    return mood_category(best)
```

An alternative method is to find the average scores for all the mood words in a mood category and choose the maximum result. This method considers all the scores for each mood category and reduces possible bias for mood categories that have a larger collection of mood words:

```
find_best_mood_category_average(word):
    (best_avg, best_cat) = (-2, "")
    foreach cat in mood_categories:
        (total, count) = (0, 0)
        foreach mood in lexicon(cat):
            total += model.dist(word, mood)
            count += 1
        cat_avg = total/count
        if cat_avg > best_avg:
            best_avg = cat_avg
            best_cat = cat
    return best_cat
```

A third option was to only calculate scores for the six mood category words and choosing the one with the highest score. This ‘label’ option explores if using a secondary collection of associated or synonym mood words was beneficial or not. Table 4 shows the results for these three options with the highest score for each in bold. Note that the mood categories are abbreviated, for example ‘sad’ instead of ‘sadness’ and so on due to space limitations.

As can be seen in Table 4, these different scoring options rarely all agree. These different methods were

explored using a test collection of messages labelled by two annotators. The test data consisted of 2,253 messages (2,062 Instagram messages and 191 tweets) collected via the AWM Stackla API. This API was used since we wanted the test data to reflect as close as possible the content expected during Remembrance Day and it had been curated as relevant by the Memorial (35 messages were irrelevant).

Table 4: Example alternative word scoring options.

Message word	Best option		Avg option		Label option	
	score	mood	score	mood	score	mood
killed	1.00	sad	0.31	sad	0.11	sad
heartfelt	0.69	resp	0.26	resp	0.49	grat
tribute	0.60	pride	0.35	resp	0.38	grat
Australian	0.11	disc	0.04	disc	0.07	resp
War	0.24	resp	0.12	resp	0.09	pride
Memorial	0.59	resp	0.23	resp	0.15	grat

These messages contained 6,374 unique words which appeared a total of 31,959 times. The three scoring options were explored by comparing the different mood allocations for each. For the 6,374 unique words processed, only 971 agreed (about 15%) for all three methods. The ‘best’ and ‘average’ methods agreed 2,339 times; ‘best’ and ‘label’: 542 times; ‘average’ and ‘label’: 952 times. Hence, the ‘average’ method had the most agreement with the other two.

A qualitative assessment was also explored by manually inspecting the words that had a different mood label, but only for the most frequently occurring words to make the task manageable. These were determined by exploring the frequency of the words in the collection of test messages: the top 300 words in terms of occurrence accounted for approximately half of all words in the test messages.

For these 300 words, 39 had the same mood label assigned by all three scoring options. For the remaining 261 words, a list was generated where the word and the two or three scoring option results were shown. Two annotators then chose the best option. Examples are shown in Table 5.

Table 5: Sample words for labelling.

Word	Mood category options
anniversary	respect, sadness
Armistice	respect, sadness
army	discontent, gratitude, pride
AWM	respect, sadness
Battalion	gratitude, respect
ceremony	respect, sadness

The annotator could also choose ‘N/A’ as the labelling result if no mood option was applicable, such as for the words ‘him’ and ‘here’. The initial annotator

agreement was 71% with 76 words resolved through discussion.

The manually labelled results were compared with the labels assigned by the three scoring options: 24% matched the ‘label’ method, 37% the ‘best’ method and 39% the ‘average’ method. These results indicate that only using the six mood words for analysis, the ‘label’ method, is not as good as using a large collection of words. Also, there is not much difference between using the maximum or average scores of the secondary words for the mood categories. Our preference was to use the average scoring option since it was considered fairer: a mood category with more secondary mood words will have an increased likelihood of having a ‘winning’ word using the best method.

We also explored the ‘N/A’ labels to see if there were characteristics that could be used to identify when words should be ignored. This is currently done with stop words, but it was hoped that the N/A words would have low scores which could be used to identify a score threshold below which words could be ignored. This was not the case.

5.2. Message Scoring Options

The process so far has used word2vec to associate a single word to a mood category. The task remains to combine these results to assign a mood to a text message. There are different methods to combine these scores to associate the message to a mood. Perhaps all the ‘winning’ mood categories could be chosen or only the mood category that is represented most. Alternatively, the highest score could be chosen as the overall result or the average for each mood category determined and the highest average chosen.

After experimenting with various options, we chose two methods for further exploration:

1. Decide on the mood category for each word in the message; sum these scores for each mood category; choose the highest sum.
2. For all words in the message calculate the average scores for each mood category; sum these scores per mood category; choose the mood category with the highest sum.

The two options are similar except that the first decides up front which mood category ‘wins’ per word, while the second defers this decision and uses the scores for each mood category for all words in the message. The first option continues the method described above where a mood category is determined for each word in the message. For the second option, the various aspects of each word in a message in terms of their similarity to the six mood categories are taken into consideration when determining the overall mood for a message. Our intuition is that including all the

information about the mood scores for all words in a message is preferable for deciding on a specific mood.

These alternatives are shown using the same example tweet text. For option 1, the results for the ‘average’ word scoring option in Table 4 can be used where the sums for the different mood categories are ‘discontent’: 0.04, ‘respect’: 0.96 and ‘sadness’ 0.31. The highest sum is for the mood ‘respect’ and so this is the mood category chosen for the message.

Option 2 requires further processing results. Tables 6 and 7 show the average scores for the six moods for each word in our example message, with the best average score highlighted in bold. These bold ‘winning’ mood categories can be compared with the results shown in Table 4 which all concur.

Table 6: Average scores per mood category.

Category	‘killed’ avg score	‘heartfelt’ avg score	‘tribute’ avg score
Discontent	0.1615	0.1494	0.1119
Gratitude	0.1285	0.2081	0.2374
Humility	0.1198	0.2171	0.1355
Pride	0.0872	0.2280	0.2805
Respect	0.1263	0.2553	0.3519
Sadness	0.3134	0.2159	0.1755

Table 7: Average scores per mood category.

Category	‘Australian’ avg score	‘War’ avg score	‘Memorial’ avg score
Discontent	0.0416	0.1109	0.0605
Gratitude	0.0137	0.1082	0.1220
Humility	0.0342	0.0532	0.0539
Pride	0.0182	0.0777	0.1447
Respect	0.0176	0.1243	0.2259
Sadness	0.0219	0.0871	0.1185

The final task of associating a mood to the message is done by summing the average scores for the six mood categories, shown in Table 8. For this example, the two methods give the same result: ‘respect’.

Table 8: Result of summing the mood categories.

Mood	Sum
Discontent	0.6358
Gratitude	0.8179
Humility	0.6137
Pride	0.8363
Respect	1.1013
Sadness	0.9323

5.3. Evaluation

The two alternative message scoring options were evaluated by comparing the results using the same test

data as before. For the 2,253 messages (2,062 Instagram posts and 191 tweets), 375 could not be processed since there are either no words in the message text found in the word2vec model or the message was empty after removal of irrelevant content.

For the remaining 1,878 messages the two message scoring options were applied, and the results were the same for 1,398 messages which is 74% agreement. The 480 messages that had different mood labels for the two methods were manually inspected by two annotators. Inspecting 480 messages was considered onerous, so only the 224 ‘published’ messages were reviewed. These are the messages vetted by AWM staff as being relevant and suitable for public display. Example messages and the mood label options are shown in Table 9.

Table 9: Sample messages for labelling.

Mood options	Message Text
respect sadness	1 poppy represents 1 life lost in WW1 there are 62000 poppy in this field. No more wars!! #ww1 #war
gratitude pride	A very Australian experience
respect sadness	Saturday afternoon in Canberra #warmemorial #mtainslie
gratitude respect	Paying our respects

The labelling of the 224 messages was undertaken by the same two annotators as before but this time there was only agreement for 129 messages (58%). The 95 differences were resolved through discussion and 97 matched the first scoring option (43%) and 127 matched the second method (57%). While this indicates the second scoring option is preferable these results are marginal. We address this in the conclusion.

6. Results

In total, for the five-week period from 5 October through to 11 November, there were almost 1,000 Facebook posts, over 1,500 tweets and more than 5,000 Instagram messages collected via the Stackla API. However, we were only interested in messages that have associated text (not all messages do), that were not published by the official Australian War Memorial social media accounts (we are interested in the mood from the general public, not the Memorial), have not been moderated by the museum staff (reflecting inappropriate or irrelevant content) and are tagged as being relevant to the Centenary of Armistice commemorations. After filtering the messages for these

extra conditions, there were 5,153 messages: 4,348 Instagram posts and 805 tweets. There are no Facebook posts since the Stackla API does not provide access to Facebook comments.

Of these, a total of 3,022 messages, about 59% included a relevant hashtag (2,274 messages), a phrase (1,353 messages) or emoji (645 messages) that could be used to associate the message to a mood using the rule conditions. Note that these rules can associate a message to multiple mood categories. By platform these numbers are 433 tweets (14%) and 2,589 Instagram messages (86%). A plot showing the number of messages per day is shown in Figure 3.

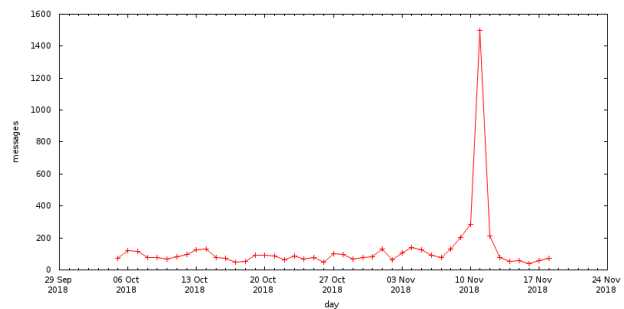


Figure 3: Number of Stackla messages per day.

The remaining messages were processed using the word2vec method described as option 2 in Section 5. In summary, for each word in the message the average cosine similarity distances, the ‘scores’, are calculated for each mood category. These averages are accumulated for all the words in the message and summed for each mood category. The mood with the highest sum is the mood associated to the message.

Note that there are still messages that cannot be processed. Of the 5,153 messages 536 (5 tweets and 531 Instagram posts) could not be associated to a mood, approximately 10%. These messages consist of a combination of URLs, user mentions, hashtags and short text messages (or no message) which when processed do not provide any information to associate the message to a mood. A summary of the different moods associated with these messages is shown in Table 10. Note that some of the messages are associated with multiple moods using the rule conditions.

Table 10: Final mood category results.

Mood	Category
Respect	2,441
Gratitude	1,729
Humility	1,466
Pride	499
Sadness	206
Discontent	195

7. Conclusions and Future Work

The social media text messages collected by the Australian War Memorial were processed to associate them to one or more of the six moods of gratitude, humility, pride, respect, sadness and discontent. This processing was done using a combination of a rule-based approach and a novel use of word embeddings with a domain specific mood lexicon.

The rules were developed in reference to a test collection of relevant social media messages. The rules associate the messages to one or more of the six mood categories based on the presence of predefined hashtags, phrases and emojis.

Different methods for determining the overall mood for a message based on the moods of the individual words were explored. The method chosen was to accumulate the average cosine similarity distances (the 'scores') for each mood found for all the words in the message and summing these averages. The mood with the highest sum is the mood associated to the message.

The word2vec processing is only performed for messages which are not assigned a mood using the rules. The rationale is that we are confident in the result of applying the rules to the messages to determine the various moods.

The processed social media content was used for the Memorial's Emotion Poppy display: an interactive, web-based visualisation showing the changing sentiment or mood expressed in the real-time social media posts discussing topics relating to the Centenary of Armistice commemorations [1]. While we have focused on mood analysis for this work, the website display was named 'Emotion Poppy' to prevent potential negative connotations relating to the use of 'mood' or 'moody' when describing people's social media messages of remembrance. During the lead up to Remembrance Day 2018, this web site showed the recently processed content while now it serves as an archive of the content collected and processed.

The use of lexicon-based approaches for sentiment or emotion analysis can suffer from issues of sparsity, when the lexicon vocabulary is infrequently present in the text, and context, when the emotive 'score' of the lexicon words are mismatched to the application domain. We have addressed both these issues through a new method to match message words to the lexicon vocabulary using the cosine similarity distance in a word2vec model and by curating our own lexicon vocabulary as a list of mood categories and associated mood words or synonyms.

These two measures allow us to successfully process a relatively small volume of social media messages and classify them into mood categories relevant to a specific topic of investigation.

There were two labelling tasks undertaken. The first, to select the best mood category produced by the three word scoring options, had good annotator agreement of 71%. The second task, to select the best mood category produced by the two message scoring options was less successful, with 57% agreement. Perhaps not unsurprisingly, this implies that the task is more successful when there is only a single word to make a judgement on. Our test sample of text messages often conveyed multiple emotions and the annotators had to choose their preferred mood with differing results. This task may have been more successful if multiple moods could be chosen as the result.

This lack of agreement in manual labelling could indicate that the mood categories and their associated words may not be as disjointed or discriminating as we assumed. We are currently investigating ways of determining this for example by using synonym lists such as WordNet [11,17].

We have used word2vec to find words that are semantically close to our lexicon words for the purposes of classifying messages into mood categories. This similarity measure can produce unintended results because words such as 'love' and 'hate' can be numerically close in the model since they can be used in similar contexts. However, in practice we have found that this issue occurs infrequently and can be mitigated by curation of the lexicon words with respect to the model being used. For our lexicon of 106 words, there were only two instances where this was an issue.

There are other areas of future work. We are currently exploring the use of different word embedding models, particularly one built using social media content. We are currently applying these techniques to see how readily they can be adapted to other issues discussed on social media. Also, the data collected can be used for supervised learning methods, for example training a text classifier. This would be useful for validating the results obtained so far.

Acknowledgements

Thanks to Aditya Joshi for the idea of using word2vec and Melissa Stubbings and Lloyd Vicencio who developed the AWM Emotion Poppy website.

8. References

- [1] Australian War Memorial Emotion Poppy <https://www.awm.gov.au/emotionpoppy/> [Accessed: 2018/11/02]
- [2] Australian War Memorial social media accounts: Facebook <https://www.facebook.com/AWMemorial/>, Instagram <https://www.instagram.com/awmemorial/>, and Twitter <https://twitter.com/AWMemorial>

- [3] Bandhakavi, A., N. Wiratunga, P. Deepak and S. Massie, "Generating a word-emotion lexicon from #emotional tweets" Proceedings of the 3rd Joint Conference on Lexical and Computational Semantics, Dublin, Ireland 2014, pp. 12-21.
- [4] Banerjee, R., K. Elgarroussi, S. Wang, Y. Zhang and C.F. Eick "Tweet Emotion Mapping: Understanding US Emotions in Time and Space" Proceedings of the 1st IEEE International Conference on Artificial Intelligence and Knowledge Engineering, Laguna Hills, CA, USA 2018, pp. 93-100.
- [5] Bhuta, S., A. Doshi, U. Doshi and M. Narvekar, "A review of techniques for sentiment analysis Of Twitter data" International Conf. on Issues and Challenges in Intelligent Computing Techniques, Ghaziabad, 2014, pp. 583-591.
- [6] Chen, Y., J. Yuan, Q. You and J. Luo, "Twitter Sentiment Analysis via Bi-sense Emoji Embedding and Attention-based LSTM" Proceedings of the 26th ACM international conference on Multimedia (MM 2018). ACM, New York, NY, USA, 117-125.
- [7] Daou, H. "Detection of Sentiment Provoking Events in Social Media" 52nd Hawaii International Conference on System Sciences (HICSS) Maui, Hawaii, USA 2019 pp 1-9.
- [8] Deshpande, M. and V. Rao, "Depression detection using emotion artificial intelligence" Proceedings of the International Conference on Intelligent Sustainable Systems, ICISS Palladam, India 2017, pp. 858-862.
- [9] Dodds, P.S., K.D. Harris, I.M. Kloumann, C.A. Bliss and C.M. Danforth, "Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter" PLoS ONE 2011 6(12).
- [10] Durahim, A.O. and M. Coşkun, "#iamhappybecause: Gross National Happiness through Twitter analysis and big data" Technological Forecasting and Social Change, Volume 99, 2015, Pages 92-105.
- [11] Fellbaum, C. (ed.) WordNet: An Electronic Lexical Database Cambridge, MA: MIT Press 1988.
- [12] Gupta, R.K., P. Bhattacharya and Y. Yang, "What constitutes happiness? Predicting and characterizing the ingredients of happiness using emotion intensity analysis" 2nd Workshop on Affective Content Analysis Honolulu, Hawaii USA 2019 pp. 152-163.
- [13] Hu, T., H. Guo, H. Sun, T.T. Nguyen and J. Luo, "Spice Up Your Chat: The Intentions and Sentiment Effects of Using Emojis" In Proceedings of the 11th International Conference on Web and Social Media, ICWSM, Montréal, Québec, Canada, May 15-18, 2017. pp 102-111
- [14] Martins, R., M. Gomes, J.J. Almeida, P. Novais and P. Henriques, "Hate speech classification in social media using emotional analysis" Proceedings of the Brazilian Conference on Intelligent Systems Sao Paulo, Brazil 2018 pp. 61-66.
- [15] Mashal, S.X. and K. Asnani, "Emotion intensity detection for social media data" Proceedings of the International Conference on Computing Methodologies and Communication, ICCMC Erode, India 2017 pp. 155-158.
- [16] Mikolov, T., K. Chen, G. Corrado and J. Dean, "Efficient Estimation of Word Representation in Vector Space" Proceedings of the International Conference on Learning Representations (ICLR) Scottsdale, Arizona USA 2013
- [17] Miller, G.A. "WordNet: A Lexical Database for English" Communications of the ACM 1995 Vol. 38, No. 11: 39-41.
- [18] Milne, D., C. Paris, H. Christensen, P. Batterham and B. O'Dea, "We Feel: Taking the emotional pulse of the world" Proceedings of the 19th Triennial Congress of the International Ergonomics Association. Melbourne, Victoria, Australia, August 2015.
- [19] Nepal, S., C. Paris, D. Georgeakopoulos, Social media for government services (2015) Springer.
- [20] Ortony, A. and T.J. Turner, "What's Basic About Emotions?" Psychological Review 1990, 97(3):315-31.
- [21] Parrott, W. "Emotions in Social Psychology" Key Readings in Social Psychology 2001. Philadelphia: Psychology Press.
- [22] Pawar, K., P. Shrishimal and R.R. Deshmukh, "Twitter Sentiment Analysis: A Review" International Journal of Scientific & Engineering Research, Volume 6, Issue 4, pp 957-964 April 2015
- [23] Rodriguez, A., C. Argueta and Y.-L. Chen, "Automatic Detection of Hate Speech on Facebook Using Sentiment and Emotion Analysis" 1st International Conference on Artificial Intelligence in Information and Communication, ICAIIC Okinawa Japan 2019, pp. 169-174.
- [24] Shaver, P., J. Schwartz, D. Kirson and C. O'Connor, "Emotional Knowledge: Further Exploration of a Prototype Approach" In G. Parrott (Eds.), Emotions in Social Psychology: Essential Readings 2001 (pp. 26-56).
- [25] Staiano, J. and M. Guerini, "DepecheMood: A lexicon for emotion analysis from crowd-annotated news" 52nd Annual Meeting of the Association for Computational Linguistics (ACL) Baltimore USA 2014 pp. 427-433.
- [26] Stieglitz, S., M. Mirbabaie and T. Potthoff, "Crisis Communication on Twitter during a Global Crisis of Volkswagen - The Case of 'Dieselgate'" 51st Hawaii International Conference on System Sciences (HICSS), Koloa, HI, USA, 2018 pp. 513-522.
- [27] Thelwall, M., K. Buckley, G. Paltoglou, D. Cai and A. Kappas, "Sentiment strength detection in short informal text" Journal of the American Society for Information Science and Technology 2010, 61(12), 2544-2558.
- [28] Thelwall, M. "Gender bias in sentiment analysis" Online Information Review 2018.
- [29] Turner, J.H. "The sociology of emotion: Basic Theoretical arguments" Emotion Review 2009. 1 (4): 340-54.
- [30] Wang, R., D. Zhou, M. Jiang, J. Si and Y. Yang, "A survey on opinion mining: From stance to product aspect" IEEE Access 2019, vol 7, pp. 41101-41124.
- [31] Zimbra, D., M. Ghiassi and S. Lee, "Brand-Related Twitter Sentiment Analysis Using Feature Engineering and the Dynamic Architecture for Artificial Neural Networks" 49th Hawaii International Conference on System Sciences (HICSS), Koloa, HI, USA, 2016 pp. 1930-1938.
- [32] Zimbra, D., A. Abbasi, D. Zeng and H. Chen, "The State-of-the-Art in Twitter Sentiment Analysis: A Review and Benchmark Evaluation" ACM Trans. Manage. Inf. Syst. 9, 2, Article 5 August 2018, 29 pages.